

การเปรียบเทียบโปรแกรมสำเร็จรูปทางสถิติในการวิเคราะห์ข้อมูล การถดถอยเชิงเส้น

A Comparison of Statistical Packages in Data Analysis of Linear Regression

ชูใจ คูหารัตนไชย^{1*}

Choojai Kuharatanachai¹

¹ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

วันที่ส่ง : 16 มกราคม 2562 วันที่แก้ไข : 30 พฤษภาคม 2562 วันที่ตอบรับ : 20 ธันวาคม 2562

บทคัดย่อ

โปรแกรมสำเร็จรูปทางสถิติ SPSS Minitab และ Microsoft Excel ถือเป็นโปรแกรมที่นิยมใช้ในการวิเคราะห์ข้อมูล ในการทำวิจัยครั้งนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์ข้อมูลการถดถอยเชิงเส้นอย่างง่าย และการถดถอยเชิงเส้นพหุคูณ โดยการจำลองข้อมูลเมื่อกำหนดค่าความแปรปรวน (σ^2) เป็นความแปรปรวนของความคลาดเคลื่อน มีค่าเท่ากับ 10, 5, 2, 1.5, 1, 0.5, 0.1 และขนาดตัวอย่าง(n) มีค่าเป็น 100, 60, 40, 20 ค่าMSE ที่คำนวณได้จากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 จะให้ค่าใกล้เคียงกันทุกสถานการณ์ที่จำลองข้อมูล และเมื่อนำค่า MSE ดังกล่าวมาทำการวิเคราะห์ความแปรปรวน จะให้ค่า p-value เข้าใกล้ 1 เกือบทุกสถานการณ์เช่นกัน

คำสำคัญ : การวิเคราะห์การถดถอย, โปรแกรมสำเร็จรูปทางสถิติ

Abstract

Statistical packages including SPSS, Minitab and Microsoft Excel are highly preferable packages for data analysis. The objective of this research is to compare IBM SPSS 22, Minitab 16 and Microsoft Excel 2007 when used in data analysis by simple linear regression and multiple linear regression methods using data simulation with the variance of error (σ^2) values of 10, 5, 2, 1.5, 1, 0.5 and 0.1 while the sample sizes are 100, 60, 40, and 20. The MSE values, calculated by the IBM SPSS 22, Minitab 16 and Microsoft Excel 2007 statistical

*ที่อยู่ติดต่อ. E-mail address : choojai.ku@kmitl.ac.th

packages, were similar in every situation of the simulation. The Analysis of Variance of such MSE values also results in p-value close to 1 in almost situation.

Keywords : Regression analysis, Statistical packages

1. บทนำ

โปรแกรมประยุกต์ทางสถิติเป็นโปรแกรมสำเร็จรูปที่ใช้กันอย่างกว้างขวางในวงการศึกษ ซึ่งบุคลากรที่ใช้งานได้แก่ นักศึกษา นักวิจัย นักสถิติ ตลอดจนนักวิชาการ ที่ต้องการคำนวณหรือวิเคราะห์หาค่าทางสถิติต่างๆ ทั้งสถิติพรรณนา และสถิติเชิงอ้างอิง เช่น การหาค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน การทดสอบค่าเฉลี่ยของประชากร 2 กลุ่มโดยใช้ t-test การทดสอบค่าเฉลี่ยของประชากรมากกว่า 2 กลุ่มโดยใช้ ANOVA การวิเคราะห์การถดถอย การแสดงกราฟ และแผนภูมิต่างๆ เป็นต้น ซึ่งโปรแกรมสำเร็จรูปทางสถิติที่นิยมใช้กัน คือ โปรแกรม R , SPSS , Minitab , SAS และ Microsoft Excel และโปรแกรมที่หาได้ง่ายและมีรูปแบบในการใช้คำสั่งที่ไม่ยุ่งยากคือ โปรแกรม SPSS , Minitab และ Microsoft Excel แต่เนื่องจากโปรแกรม Microsoft Excel จะเป็นโปรแกรมที่หาได้ง่ายและมีค่าใช้จ่ายน้อยกว่าเมื่อเทียบกับโปรแกรม SPSS , Minitab และการวิเคราะห์การถดถอย(Regression analysis) เป็นวิธีหนึ่งที่นำมาใช้มากในการวิเคราะห์ทางด้านสถิติอยู่ตลอดเวลา ซึ่งวิธีนี้เป็นการศึกษาถึงความสัมพันธ์ของตัวแปรตั้งแต่ 2 ตัวขึ้นไป โดยมีวัตถุประสงค์ที่จะประมาณหรือพยากรณ์ค่าของตัวแปรตัวหนึ่ง ในที่นี้จะให้เป็นตัวแปรตาม Y จากตัวแปรอื่นๆ หรือตัวแปรอิสระ $X_1, X_2, \dots, X_k$ ที่มีความสัมพันธ์กับตัวแปรที่ต้องการพยากรณ์ ถ้าตัวแปรตามกับตัวแปรอิสระมีความสัมพันธ์ในรูปแบบเส้นตรง จะเรียกว่า การถดถอยเชิงเส้น (Linear regression) ซึ่งสามารถแบ่งเป็น 2 แบบ คือ การถดถอยเชิงเส้นอย่างง่าย (Simple linear regression) เป็นการวิเคราะห์การถดถอยของตัวแปรอิสระ 1 ตัวและตัวแปรตาม 1 ตัว โดยตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรง และการถดถอยเชิงเส้นพหุคูณ (Multiple linear regression) เป็นการวิเคราะห์การถดถอยของตัวแปรอิสระมากกว่า 1 ตัวและตัวแปรตาม 1 ตัว โดยตัวแปรอิสระกับตัวแปรตามมีความสัมพันธ์เชิงเส้นตรง

จากงานวิจัยในการเปรียบเทียบความน่าเชื่อถือของโปรแกรมสำเร็จรูปสำหรับงานวิเคราะห์ทางสถิติ ได้ทำการประมาณค่าทางสถิติ 3 วิธีคือ การวิเคราะห์สถิติตัวแปรเดียว (Univariate Summary Statistic), การวิเคราะห์ความแปรปรวนจำแนกทางเดียว (One-Way Analysis of Variance) และการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) ของโปรแกรมสำเร็จรูป 5 โปรแกรม ประกอบด้วยโปรแกรม SAS 9, R 2.6.2, Microsoft Excel 2007, SPSS 15 และ Minitab 15 โดยใช้ข้อมูลจาก The National Institute of Standards and Technology (NIST) ซึ่งแบ่งข้อมูลตามระดับความยาก 3 ระดับ คือ ระดับต่ำ, ระดับปานกลาง และระดับสูง ซึ่งผลการวิจัยสรุปในส่วนของการวิเคราะห์การถดถอยพบว่า ค่าความแม่นยำของสถิติทดสอบเอฟของโปรแกรม SAS 9, R 2.6.2 และ Microsoft Excel 2007 มีค่าความแม่นยำอยู่ในระดับสูง ส่วนโปรแกรม SPSS 15 และ Minitab 15 เมื่อข้อมูลมีความยากระดับสูง จะให้ค่า

ความแม่นยำในระดับต่ำ ส่วนค่าความแม่นยำของค่าสัมประสิทธิ์การถดถอย เมื่อข้อมูลมีความยากระดับต่ำ และปานกลาง ทั้ง 5 โปรแกรม จะให้ค่าความแม่นยำในระดับสูง แต่เมื่อข้อมูลมีความยากระดับสูง พบว่า โปรแกรม SAS 9 , R 2.6.2 , Microsoft Excel 2007 และ Minitab 15 ให้ค่าความแม่นยำไม่แตกต่างกัน แต่ โปรแกรม SPSS 15 ให้ค่าความแม่นยำต่ำกว่าโปรแกรมอื่น [1] และจากการศึกษางานวิจัย การเปรียบเทียบ การใช้โปรแกรม Minitab, Instat และ R Commander กรณีการทดสอบด้วยไคกำลังสอง โดยใช้โจทย์ ปัญหา 3 ตัวอย่าง มาทำการทดสอบด้วยไคกำลังสอง พบว่า การคำนวณค่าทางสถิติของการทดสอบด้วยไค กำลังสองจากทั้ง 3 โปรแกรม ให้ผลลัพธ์ในข้อสรุปที่เหมือนกัน เพียงแต่มีความแตกต่างกันในเรื่องการ นำเข้าข้อมูลและการแจ้งข้อความเตือนเมื่อความถี่คาดหวังมีค่าน้อยกว่า 5 เท่านั้น [2]

จากที่กล่าวมาจะเห็นได้ว่า ในการวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปทางสถิติที่แตกต่างกัน จะ ให้ผลลัพธ์ทางสถิติในเรื่องเดียวกันที่เหมือนหรือแตกต่างกันก็ได้ และจากการวิเคราะห์การถดถอยเชิงเส้น อย่างง่ายและเชิงเส้นพหุคูณ สามารถใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS Minitab และ Microsoft Excel มาช่วยในการวิเคราะห์ข้อมูล ผู้วิจัยจึงเห็นถึงความสำคัญในการเปรียบเทียบโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์การถดถอยเชิงเส้น เพื่อให้ผู้ใช้ โปรแกรมสำเร็จรูปทางสถิติสามารถเลือกใช้โปรแกรมได้อย่างเหมาะสมกับลักษณะของข้อมูลที่ใช้ในการ วิเคราะห์

2. วัตถุประสงค์

เพื่อเปรียบเทียบโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์ข้อมูลการถดถอยเชิงเส้นอย่างง่ายและการถดถอยเชิงเส้นพหุคูณ

3. การดำเนินงานวิจัย

3.1 ศึกษารูปแบบการถดถอยเชิงเส้น

- (ก) รูปแบบการถดถอยเชิงเส้นอย่างง่าย เป็นรูปแบบกรณีที่ตัวแปรตาม Y มีความสัมพันธ์เชิงสถิติกับตัวแปรอิสระ X เพียงตัวแปรเดียว และลักษณะความสัมพันธ์เป็นแบบเส้นตรง

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- (ข) รูปแบบการถดถอยเชิงเส้นพหุคูณ เป็นรูปแบบกรณีที่ตัวแปรตาม Y มีความสัมพันธ์เชิงสถิติกับตัวแปรอิสระ X มากกว่าหนึ่งตัว และลักษณะความสัมพันธ์เป็นแบบเส้นตรง ในที่นี้จะศึกษาเฉพาะตัวแปรอิสระ X_1 และ X_2

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

เมื่อ β_0 คือ ส่วนตัดแกน Y

β_1 , β_2 คือ สัมประสิทธิ์การถดถอย

โดยมีข้อสมมติของรูปแบบการถดถอยดังนี้

1. ค่าตัวแปร X หรือตัวแปร X_1, X_2 จะต้องเป็นค่าที่กำหนดไว้ล่วงหน้าหรือทราบค่า
2. ความคลาดเคลื่อน ε มีการแจกแจงแบบปกติมีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเท่ากับ σ^2 นั่นคือ $\varepsilon \sim N(0, \sigma^2)$
3. ε_i และ ε_j สำหรับ $i \neq j$ มีการแจกแจงที่เป็นอิสระกัน จะทำให้ $\text{COV}(\varepsilon_i, \varepsilon_j) = 0$ [3]

3.2 การจำลองข้อมูล

ในการจำลองครั้งนี้ จะแบ่งออกเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 จะทำการเปรียบเทียบโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์ข้อมูลการถดถอยเชิงเส้นอย่างง่ายจากรูปแบบ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ กำหนดให้ $\beta_0 = 5, \beta_1 = 1, X \sim N(3, 2.25)$ และ $\varepsilon \sim N(0, \sigma^2)$ เมื่อ σ^2 มีค่าเท่ากับ 10, 5, 2, 1.5, 1, 0.5 และ 0.1 สำหรับขนาดตัวอย่าง(n) คือ 100, 60, 40 และ 20 โดยมีขั้นตอนดังนี้

1. จากรูปแบบการถดถอยเชิงเส้นอย่างง่าย $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ จะทำการจำลองข้อมูลของตัวแปรอิสระ $X \sim N(3, 2.25)$ และค่าคลาดเคลื่อน $\varepsilon \sim N(0, \sigma^2)$ จำนวน n ค่าในการสร้างค่า Y_i จำนวน n ค่า
2. นำค่า X_i และค่า Y_i ที่ได้มาวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย เพื่อหาค่าความคลาดเคลื่อนกำลังสอง (Mean Square Error ; MSE) โดยใช้โปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007
3. ทำการจำลองแต่ละสถานการณ์จำนวน 100 ครั้ง
4. ในแต่ละสถานการณ์ จะทำการเปรียบเทียบค่า MSE ที่ได้จากโปรแกรมทั้ง 3 ด้วยการวิเคราะห์ความแปรปรวน (Analysis of Variance) หรือ การทดสอบความแปรปรวนด้วยอันดับครัสคัลและวอลลิส (the Kruskal-Wallis One-Way Analysis of Variance by Ranks)

ส่วนที่ 2 จะทำการเปรียบเทียบโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์ข้อมูลการถดถอยเชิงเส้นพหุคูณ จากรูปแบบ $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ กำหนดให้ $\beta_0 = 5, \beta_1 = 1, \beta_2 = 3, X_1 \sim N(3, 2.25), X_2 \sim N(5, 4)$ และ $\varepsilon \sim N(0, \sigma^2)$ เมื่อ σ^2 มีค่าเท่ากับ 10, 5, 2, 1.5, 1, 0.5 และ 0.1 สำหรับขนาดตัวอย่าง(n) คือ 100, 60, 40 และ 20 โดยมีขั้นตอนดังนี้

1. จากรูปแบบการถดถอยเชิงเส้นพหุคูณ $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ จะทำการจำลองข้อมูลของตัวแปรอิสระ $X_1 \sim N(3, 2.25), X_2 \sim N(5, 4)$ และค่าคลาดเคลื่อน $\varepsilon \sim N(0, \sigma^2)$ จำนวน n ค่าในการสร้างค่า Y_i จำนวน n ค่า

- นำค่า X_{1i} , X_{2i} และค่า Y_i ที่ได้มาวิเคราะห์การถดถอยเชิงเส้นพหุคูณ เพื่อหาค่าความคลาดเคลื่อนกำลังสอง (Mean Square Error ; MSE) โดยใช้โปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007
- ทำการจำลองแต่ละสถานการณ์จำนวน 100 ครั้ง
- ในแต่ละสถานการณ์ จะทำการเปรียบเทียบค่า MSE ที่ได้จากโปรแกรมทั้ง 3 ด้วยการวิเคราะห์ความแปรปรวน (Analysis of Variance) หรือ การทดสอบความแปรปรวนด้วยอันดับครัสคัลและวอลลิส (the Kruskal-Wallis One-Way Analysis of Variance by Ranks)

4. ผลการวิเคราะห์

จากการจำลองข้อมูลตัวแปรอิสระ (X) และตัวแปรตาม (Y) มาวิเคราะห์สมการถดถอย เพื่อหาค่าความคลาดเคลื่อนกำลังสอง (MSE) โดยใช้โปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 ในแต่ละสถานการณ์จำนวน 100 ครั้ง ค่า MSE ที่ได้จากทั้ง 3 โปรแกรมจะนำมาทดสอบเกี่ยวกับข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนก่อน แล้วจึงนำมาเปรียบเทียบด้วยการวิเคราะห์ความแปรปรวน (ANOVA) หรือการทดสอบความแปรปรวนด้วยอันดับครัสคัลและวอลลิส ได้ผลดังนี้

ตารางที่ 1. แสดงผลการทดสอบค่าเฉลี่ยของค่า MSE ที่คำนวณจากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22 , Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย

สถานการณ์จำลอง		ค่าเฉลี่ยของค่า MSE ที่คำนวณได้จากโปรแกรม			ค่าสถิติ		p-value
ค่าความแปรปรวน (σ^2)	ขนาดตัวอย่าง (n)	SPSS 22	Minitab 16	Excel 2007	ค่า F	ค่า H	
10	100	10.11649	10.11662	10.11166	-	0.00	1.00
	60	10.08839	10.09860	10.08840	0.00122	-	0.99877
	40	10.06692	10.06538	10.06694	0.00001	-	0.99998
	20	10.29585	10.29590	10.29590	0.00	-	1.00
5	100	5.32990	5.33025	5.32983	-	0.065	0.968
	60	5.41450	5.41465	5.41454	-	0.004	0.998
	40	5.36350	5.36385	5.36352	-	0.003	0.999
	20	4.99750	4.99750	4.99752	-	0.020	0.990
2	100	2.07230	2.07150	2.07234	-	0.138	0.934
	60	2.10360	2.10330	2.10355	-	0.005	0.997
	40	2.15405	2.15390	2.15417	-	0.061	0.970
	20	2.14480	2.14480	2.14470	-	0.045	0.978

ตารางที่ 1. (ต่อ)

สถานการณ์จำลอง		ค่าเฉลี่ยของค่า MSE ที่คำนวณได้จากโปรแกรม			ค่าสถิติ		p-value
ค่าความแปรปรวน (σ^2)	ขนาดตัวอย่าง (n)	SPSS 22	Minitab 16	Excel 2007	ค่า F	ค่า H	
1.5	100	1.42540	1.42600	1.42542	-	0.047	0.977
	60	1.47395	1.47275	1.47391	-	0.180	0.914
	40	1.47575	1.47595	1.47579	-	0.012	0.994
	20	1.57555	1.57555	1.57550	-	0.020	0.990
1	100	1.00344	1.00370	1.0034	0.00013	-	0.99986
	60	1.01474	1.01466	1.01472	0.00	-	0.99999
	40	1.02193	1.02175	1.02195	0.00002	-	0.99997
	20	1.01873	1.01874	1.01879	0.00	-	1.00
0.5	100	0.49085	0.49110	0.49083	-	0.062	0.969
	60	0.47532	0.47610	0.47534	-	0.227	0.893
	40	0.45483	0.45474	0.45479	-	0.006	0.997
	20	0.43806	0.43806	0.43798	-	0.009	0.996
0.1	100	0.11066	0.11120	0.11073	-	0.123	0.940
	60	0.11771	0.11726	0.11771	-	0.012	0.994
	40	0.10237	0.10237	0.10237	0.00	-	1.00
	20	0.10208	0.10208	0.10206	0.00	-	1.00

จากตารางที่ 1. จะพบว่า ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายด้วยการจำลองข้อมูล เมื่อกำหนดค่าความแปรปรวน (σ^2) มีค่าเท่ากับ 10, 5, 2, 1.5, 1, 0.5, 0.1 และขนาดตัวอย่าง (n) มีค่าเป็น 100, 60, 40, 20 ค่าเฉลี่ยของค่า MSE ที่คำนวณได้จากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 จะให้ค่าใกล้เคียงกันทุกสถานการณ์ที่จำลองข้อมูล และเมื่อนำค่า MSE ดังกล่าวมาทำการวิเคราะห์ความแปรปรวน (ANOVA) หรือทดสอบความแปรปรวนด้วยอันดับครัสคัลและวอลลิส จะให้ค่า p-value เข้าใกล้ 1 ทุกสถานการณ์เช่นกัน

ตารางที่ 2. แสดงผลการทดสอบค่าเฉลี่ยของค่า MSE ที่คำนวณจากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22 , Minitab 16 และ Microsoft Excel 2007 ในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

สถานการณ์จำลอง		ค่าเฉลี่ยของค่า MSE ที่คำนวณได้จากโปรแกรม			ค่าสถิติ		p-value
ค่าความแปรปรวน (σ^2)	ขนาดตัวอย่าง (n)						
		SPSS 22	Minitab 16	Excel 2007	ค่า F	ค่า H	
10	100	10.10041	10.09411	10.10045	-	0.010	0.995
	60	10.04161	10.04080	10.04159	0.00001	-	0.99999
	40	10.03807	10.03780	10.03812	0.00	-	1.00
	20	10.21130	10.21198	10.21136	0.00	-	1.00
5	100	5.33260	5.34000	5.33263	-	0.180	0.914
	60	5.432750	5.42450	5.43272	-	0.321	0.852
	40	5.35180	5.34800	5.35175	-	0.005	0.998
	20	4.98830	4.98900	4.98830	-	0.022	0.989
2	100	2.02910	2.08500	2.07911	-	1.828	0.401
	60	2.11665	2.12150	2.11676	-	0.155	0.925
	40	2.14475	2.14750	2.14484	-	0.060	0.971
	20	2.06930	2.06980	2.06936	-	0.043	0.979
1.5	100	1.42470	1.42000	1.42473	-	0.097	0.953
	60	1.46750	1.46800	1.46748	-	0.055	0.973
	40	1.47090	1.46500	1.47084	-	0.240	0.887
	20	1.57025	1.56950	1.57029	-	0.003	0.999
1	100	1.00381	1.00200	1.00384	-	0.002	0.999
	60	1.01559	1.01780	1.01534	-	0.003	0.999
	40	1.02090	1.02040	1.02094	0.00019	-	0.99981
	20	1.01567	1.01612	1.01565	0.00006	-	0.99994
0.5	100	0.48810	0.48700	0.48819	-	0.866	0.648
	60	0.47210	0.47000	0.47207	-	0.620	0.733
	40	0.44962	0.45460	0.44970	-	0.237	0.888
	20	0.41510	0.41410	0.41500	-	0.052	0.974
0.1	100	0.10162	0.10185	0.10162	-	0.325	0.850
	60	0.10244	0.10279	0.10244	-	0.165	0.921
	40	0.10230	0.10222	0.10229	-	0.164	0.921
	20	0.10206	0.10199	0.10206	-	0.015	0.992

จากตารางที่ 2 จะพบว่า ในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณด้วยการจำลองข้อมูลเมื่อ กำหนดค่าความแปรปรวน (σ^2) มีค่าเท่ากับ 10, 5, 2, 1.5, 1, 0.5, 0.1 และขนาดตัวอย่าง(n) มีค่าเป็น 100, 60, 40, 20 ค่าเฉลี่ยของค่าMSE ที่คำนวณได้จากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 จะให้ค่าใกล้เคียงกันทุกสถานการณ์ที่จำลองข้อมูล และเมื่อนำ ค่า MSE ดังกล่าวมาทำการวิเคราะห์ความแปรปรวน (ANOVA) หรือทดสอบความแปรปรวนด้วยอันดับครัสคัลและวอลลิส จะให้ค่า p-value เข้าใกล้ 1 เกือบทุกสถานการณ์ ยกเว้นกรณีที่ $\sigma^2 = 2$ และ $\sigma^2 = 0.5$ เมื่อ n = 100 ได้ค่า p-value = 0.401 และ 0.648 ตามลำดับ

5. สรุปผลการวิเคราะห์

จากการจำลองข้อมูลเพื่อใช้ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายและการถดถอยเชิงเส้นพหุคูณ จะเห็นว่า ค่าMSE ที่คำนวณได้จากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 จะให้ค่าใกล้เคียงกับค่าความแปรปรวน (σ^2) ที่กำหนดในการจำลองข้อมูลในทุกๆ ขนาดตัวอย่าง และเมื่อกำหนดค่าความแปรปรวน (σ^2) และขนาดตัวอย่าง(n) มีค่าที่แตกต่างกันในการจำลองแต่ละสถานการณ์ จะเห็นว่า ค่าMSE ที่คำนวณได้จากโปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 จะให้ค่าใกล้เคียงกันทุกสถานการณ์ที่จำลองข้อมูล และเมื่อนำ ค่า MSE ดังกล่าวมาทำการวิเคราะห์ความแปรปรวน (ANOVA) หรือทดสอบความแปรปรวนด้วยอันดับครัสคัลและวอลลิส จะให้ค่า p-value เข้าใกล้ 1 ทุกสถานการณ์ แสดงว่า ในการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายและการถดถอยเชิงเส้นพหุคูณสามารถใช้โปรแกรมสำเร็จรูปทางสถิติ IBM SPSS 22, Minitab 16 และ Microsoft Excel 2007 มาช่วยในการวิเคราะห์ข้อมูลได้ โดยผลลัพธ์ที่ได้จากโปรแกรมทั้ง 3 จะให้ผลที่ไม่แตกต่างกัน

เอกสารอ้างอิง (References)

- [1] กมล บุชบา สายทอง อมรวิเชษฐ์ และ สารตริย์ วัชรารณณ์. 2552. การศึกษาเปรียบเทียบความน่าเชื่อถือของโปรแกรมสำเร็จรูปสำหรับงานวิเคราะห์ทางสถิติ. กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์. [Kamao Budsaba Saitong Amornwichee and Saratrai Watcharaporn. 2009. A Comparative Study on the Reliability of Software Packages for Statistical Analysis. Bangkok: Thammasat University. (in Thai)]
- [2] ปรีดาภรณ์ กาญจนสำราญวงศ์. 2555. การเปรียบเทียบการใช้โปรแกรม Minitab, Instat และ R Commander กรณีการทดสอบด้วยไคกำลังสอง. สงขลา: มหาวิทยาลัยทักษิณ. [Preedaporn Kanjanasamranwong. 2012. The Comparison of using Minitab, Instat and R Commander for Chi-square Test. Songkhla: Thaksin University. (in Thai)]

- [3] ทรงศิริ แท้สมบัติ. 2542. การวิเคราะห์การถดถอย. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์. [Tsongiri Taesombat. 1999. Regression Analysis. Bangkok: Kasetsart University Publishing Company. (in Thai)]